Technical report

Embedded Reduced Models in Flowsheet Optimization

Yidong Lang and Lorenz T. Biegler[+]

Department of Chemical Engineering

Carnegie Mellon University

Pittsburgh, PA 15213

Abstract

Advanced energy systems demand powerful and systematic optimization strategies for analysis, high performance design and efficient operation. Such processes are modeled through a heterogeneous collection of device-scale and process scale models, which contain distributed and lumped parameter models of varying complexity. This work addresses the integration and optimization of advanced energy models through multi-scale optimization strategies. In particular, we consider the optimal design of advanced energy processes by merging device-scale (e.g., CFD) models with flowsheet simulation models through sophisticated model reduction strategies. Recent developments in surrogate-based optimization have led to a general decomposition framework with multiple scales and convergence guarantees to the overall multi-scale optimum. Here, we develop two trust region-based algorithms where gradients are not required from the original detailed model (ODM). These algorithms borrow from derivative-free optimization (DFO) methods for unconstrained optimization and we extend them to the constrained case, as well as to process flowsheet optimization with different scale models that allow only limited recourse to the ODM. Both methods demonstrate multi-scale optimization of advanced energy processes. The resulting theoretical derivations and developed algorithms are interesting and justify our previous work, including methodologies of reduced model (RM) development and flowsheet optimization, with reduced models based on their CFD counterparts.

1 Introduction

With few exceptions, process simulation models consist of lumped parameter descriptions with a number of ideal assumptions (e.g., perfect mixing, plug flow, equilibrium behavior and shortcut models). On the other hand, there is a growing need for more detailed, multi-scale process models, particularly for process design and optimization. As an example, consider the IGCC process flowsheet in Figure 1 with three units: the gasifier, temperature swing adsorber (TSA) and the combustor, which cannot be modeled accurately with conventional process simulation models. The process model comprises lumped parameter (algebraic) models for heat exchange and compression, distributed, dynamic (differential-algebraic) models for the TSA unit and distributed, multi-phase (partial differential-algebraic) CFD models for the gasification and combustion units. To provide the optimization capability for the overall process model, our strategy replaces these heterogeneous (DAE, PDAE) models by reduced models (RMs) consisting of algebraic equations, and applies large-scale equation-based optimization strategies. In this way we hope to capture the phenomena of multi-phase flow, particle mechanics, and dynamic operation within the process optimization.

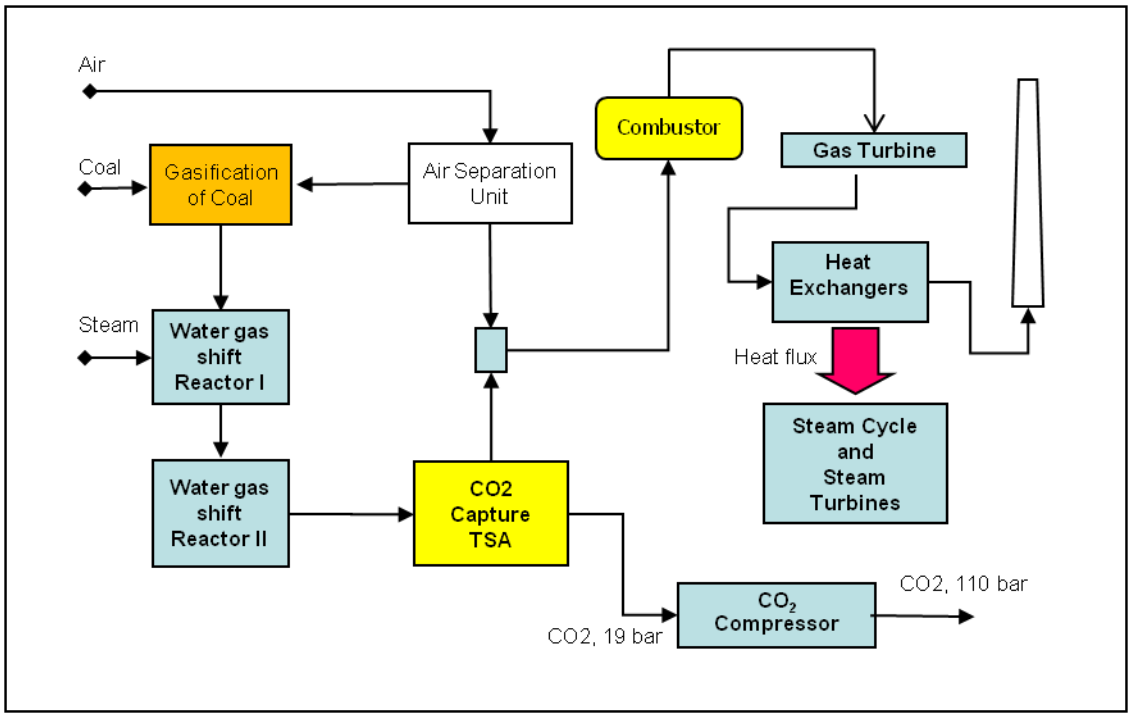


Figure 1. IGCC Flowsheet with Carbon Capture

In our previous technical report, we considered the construction of reduced models (RM) within the process flowsheet. These RMs balance accuracy with computational cost. We now develop and analyze an optimization framework with reduced models that leads to convergence to the optimum of the original system models. A key question is whether this convergent optimization framework can be performed without frequent recourse to the original models. We answer this question in the following sections. In the next section we describe the DFO trust region algorithm for unconstrained optimization as developed in Conn et al. (2009). This algorithm and its associated convergence analysis form the basis for our multi-scale flowsheet

optimization strategy. Section 3 then provides a general description of the RMs along with the *fully linear* property for optimization. We also formulate two flowsheet optimization problems, one with original detailed models (ODMs) and the second with RMs, and prove that if the *fully linear* property holds for the RMs, it holds for the *entire* RM-based flowsheet model as well. This allows us to reformulate the flowsheet optimization problem as an inequality constrained problem only in the space of the independent variables. Section 5 then develops a trust region algorithm for the reformulated problem based on the DFO algorithm described in Section 2. Finally, Section 6 considers special cases of this algorithm that lead to greater efficiency of the RM-based strategy. In addition, several examples are presented that describe this approach.

2 The Basic Trust Region Algorithm

Consider the unconstrained optimization problem given by:

$$\min_{x} f(x) \tag{1}$$

and the corresponding trust region problem solved at x_k based on a reduced model:

$$\min_{x} f^r(x_k + s) \tag{2}$$

Within the trust region Δ_k it is essential that $f^r(x)$ approximates $f(x)$ well. To do this we define the fully linear property where:

$$| f(x) - f^r(x) | \le \kappa_f \Delta^2, \quad \| \nabla f(x) - \nabla f^r(x) | \le \kappa_g \Delta \tag{3}$$

The unconstrained trust region algorithm of Conn et al. (2008) can be summarized briefly as follows.

Algorithm I

1. Choose constants $0 \le \eta_0 \le \eta_1 < 1$ (with $\eta_1 > 0$), $\gamma \in (0,1)$, $\gamma_{inc} > 1$ tolerance ε_t and $\mu > 0$. Choose an initial point x_0 and $\Delta_{\max} > 0$ along with Δ_0 and model $f^r(x_0 + s)$.

2. At iteration k, if $\left\| \nabla f^r(x_k) \right\| > \varepsilon_t$ then go to Step 3. Else, if $\Delta_k > \mu \| \nabla f^r(x_k) \|$ or $f^r(x_k + s)$ is not fully linear for $s \in \Delta_k$, then construct a new model $f^r(x_k + s)$ that is fully linear on a (possibly smaller) trust region with $\Delta_k := \min(\mu \| \nabla f^r(x_k) \|, \Delta_k)$.

3. Solve (approximately) the trust region problem (2) and compute the step s_k that provides a fraction of Cauchy decrease [1], i.e.,

$$f^r(x_k) - f^r(x_k + s_k) \ge \frac{\kappa_{fcd}}{2} \| \nabla f^r(x_k) \| \Delta_k \tag{4}$$

4. Assess the acceptability of the trial point; define

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{f^r(x_k) - f^r(x_k + s_k)} \tag{5}$$

and consider the following conditions:

- If $\rho_k \ge \eta_1$, set $x_{k+1} = x_k + s_k$ and potentially increase trust region using $\Delta_{k+1} = \min(\gamma_{inc} \Delta_k, \Delta_{\max})$. This is called a *successful iteration*.
- If $\rho_k \in [\eta_0, \eta_1)$ and the model is fully linear, set $x_{k+1} = x_k + s_k$ and decrease trust

region, $\Delta_{k+1} = \gamma \cdot \Delta_k$ .This is called an *acceptable iteration*.

- If $\rho_k < \eta_1$ and the model is not fully linear, then set $x_{k+1} = x_k$ and apply model refinement to attempt to make $f^r(x)$ fully linear on Δ_k. This is called a *model improving iteration*.
- If $\rho_k < \eta_0$ and the model is fully linear, then set $x_{k+1} = x_k$ and decrease trust region, $\Delta_{k+1} = \gamma \cdot \Delta_k$. This is called an *unsuccessful iteration*.

5. Set $k = k + 1$ and go to step 2.

Under the assumptions that $f(x)$ is bounded below, that the fully linear property holds (or can be made to hold) at x_k and that the Hessian has a bounded norm, Conn et al. (2008) proved that $\lim_{k\to\infty} \nabla f(x_k) = 0$. Algorithm I and its properties therefore lead to a powerful result: convergence to a stationary point of $f(x)$ does not require direct evaluation of $\nabla f(x)$. Instead, only $\nabla f^r(x)$ is needed. As seen above, convergence is enabled by requiring fully linear models as well as $\lim_{k\to\infty} \Delta_k = 0$.

For multiscale flowsheet optimization with reduced models, we extend this unconstrained algorithm to include two important features. First, in the next section we show that reduced flowsheet models with the fully linear property, can be embedded within the flowsheet, and this leads to objective and constraint functions for the *overall flowsheet model* that also share the fully linear property. Second, in Section 4, we extend Algorithm I to deal with inequality constraints through minimization of an l_1 penalty function, which is equivalent to the RM-based flowsheet optimization problem.

3 Formulation for Integrated RM-based Process Optimization

As seen in the flowsheet in Figure 1, we can describe the ODMs in the flowsheet by the following equations:

$$y - \bar{d}(w,u) = 0, \quad \hat{d}(w,u) = 0 \tag{6}$$

where u is the vector of model inputs from the rest of the process flowsheet (including decision variables for optimization), y is the vector of model outputs to the rest of the process flowsheet (including response variables for optimization) and w is the vector of internal (dependent) variables within the ODM. Similarly we describe the RM by:

$$y - \bar{r}(w_r,u) = 0, \quad \hat{r}(w_r,u) = 0 \tag{7}$$

where w_r is the vector of internal variables within the RM. We assume that both $\hat{r}(w_r,u) = 0$ and $\hat{d}(w,u) = 0$ are regular systems of equations where the implicit function theorem can be applied to eliminate the variables w_r and w and rewrite the RM and ODM as $y = r(u)$ and $y = d(u)$, respectively. Moreover, we assume that the following properties are satisfied:

- The RM is constructed from an experimental design (with points in set K) and satisfies interpolation properties: $y = r(u_k) = d(u_k)$, $k \in K$.
- The ODM and RM are at least twice differentiable and their first derivatives are Lipschitz continuous over the entire domain of u.

[1] In Conn et al. (2008), a quadratic term was also introduced in (4). Here we consider the more restrictive case where the Hessian term is ignored. This does not affect the first order convergence proof in Conn et al (2008).

- The *fully linear* property holds between the ODM and RM, i.e.:

$$\| d(u) - r(u) \| \le \kappa_f \Delta^2, \qquad \| \nabla d(u) - \nabla r(u) \| \le \kappa_g \Delta \tag{8}$$

The flowsheet optimization problem with embedded ODMs can be written as:

$$\min \phi(u, z, y), \quad s.t. \;\; h(u, z, y) \le 0, \quad c(u, z, y) = 0, \quad y - d(u) = 0 \tag{9}$$

where z is the vector of additional flowsheeting variables and the functions for the scalar objective, inequality constraints and additional equations for the flowsheet, $\phi(u, z, y)$, $h(u, z, y)$ and $c(u, z, y)$, respectively, are assumed to be twice differentiable. Similarly, the flowsheet optimization problem with embedded RMs can be written as:

$$\min \phi(u, z, y), \quad s.t. \;\; h(u, z, y) \le 0, \quad c(u, z, y) = 0, \quad y - r(u) = 0 \tag{10}$$

For the development of the RM-based flowsheet optimization strategy, we need the flowsheet models for (9) and (10) to satisfy the fully linear property as well. To show this we partition the variables (u, z, y) into (x, v) where x and v, respectively, are the independent and dependent variables with respect to the equality constraints, $c(u, z, y) = 0$. From the structure of the ODMs and RMs we see that y can also be classified as a dependent variable vector and eliminated directly, while the vectors u and z may need to be partitioned into independent and dependent subvectors, i.e., (u_x, z_x) and (u_v, z_v), respectively. We also assume that the flowsheet equations

$$c(u, z, d(u)) = c_d(x, v) = 0$$

are regular and the implicit function theorem can be applied to form: $v = v_d(x)$. Similarly, the equations with RMs can be written as

$$c(u, z, r(u)) = c_r(x, v) = 0.$$

These equations are also regular and the implicit function theorem can be applied to form: $v = v_r(x)$.

Moreover, for the constraints in (9) and (10) we note that for a fixed value of x, v_r and v_d will take different values, i.e.,

$$c(u_x, u_v, z_x, z_v, d(u_x, u_v)) = c(x, v_d, d(x, v_d)) = c_d(x, v_d) = 0 \tag{11}$$

$$c(u_x, u_v, z_x, z_v, r(u_x, u_v)) = c(x, v_r, r(x, v_r)) = c_r(x, v_r) = 0 \tag{12}$$

Equating the two relations leads to:

$$\begin{aligned} 0 = c(x, v_r, r(x, v_r)) = c(x, v_d, d(x, v_d)) &+ [c(x, v_r, d(x, v_r)) - c(x, v_d, d(x, v_d))] \\ &+ [c(x, v_r, r(x, v_r)) - c(x, v_r, d(x, v_r))] \end{aligned} \tag{13}$$

or, equivalently:

$$0 = c_r(x, v_r) = c_d(x, v_d) + [c_d(x, v_r) - c_d(x, v_d)] + [c_r(x, v_r) - c_d(x, v_r)] \tag{14}$$

and substituting for the bracketed terms leads to:

$$0 = A_v(v_r - v_d) + A_y(r(x,v_r) - d(x,v_r)) \tag{15}$$

From the mean value theorem,

$$A_v = \int_0^1 \nabla_v c_d(x, v(\tau))^T d\tau \quad \text{for} \quad v(\tau) = v_d + \tau(v_r - v_d)$$

$$A_y = \int_0^1 \nabla_y c(x, v_r, y(\tau))^T d\tau \quad \text{for} \quad y(\tau) = d(x,v_r) + \tau(r(x,v_r) - d(x,v_r))$$

From regularity of the equalities, we know that A_v is square and nonsingular. As a result we have:

$$(v_r - v_d) = -(A_v)^{-1} A_y(r(x,v_r) - d(x,v_r)) \tag{16}$$

$$\Rightarrow \| v_r - v_d \| \le C_2 \| r(x,v_r) - d(x,v_r) \| \le C_2 \Delta^2 \tag{17}$$

where the last inequality follows from the fully linear property.

For (9) and (10) we now consider the objective function and inequalities as follows:

$$\begin{aligned}
\phi(u_x, u_v, z_x, z_v, d(u_x, u_v)) &= \phi_d(x,v) \\
\phi(u_x, u_v, z_x, z_v, r(u_x, u_v)) &= \phi_r(x,v) \\
h(u_x, u_v, z_x, z_v, d(u_x, u_v)) &= h_d(x,v) \\
h(u_x, u_v, z_x, z_v, r(u_x, u_v)) &= h_r(x,v)
\end{aligned}$$

At points where the equalities are satisfied we define these functions as follows. For the ODM we have $c_d(x, v_d) = 0$ and also:

$$\phi_d(x, v_d) = f(x) \text{ and } h_d(x, v_d) = g(x)$$

and also for the RM we have $c_r(x, v_r) = 0$

$$\phi_r(x, v_r) = f^r(x) \text{ and } h_r(x, v_r) = g^r(x)$$

We now can write the following set of reduced gradients:

$$\frac{d\phi_d(x,v_d)}{dx} = \frac{\partial \phi_d(x,v_d)}{\partial x} + \frac{\partial d(x,v_d)}{\partial x}\frac{\partial \phi_d(x,v_d)}{\partial y} \tag{18}$$

$$\frac{d\phi_d(x,v_d)}{dv} = \frac{\partial \phi_d(x,v_d)}{\partial v} + \frac{\partial d(x,v_d)}{\partial v}\frac{\partial \phi_d(x,v_d)}{\partial y} \tag{19}$$

$$\frac{d\phi_r(x,v_r)}{dx} = \frac{\partial \phi_r(x,v_r)}{\partial x} + \frac{\partial r(x,v_r)}{\partial x}\frac{\partial \phi_r(x,v_r)}{\partial y} \tag{20}$$

$$\frac{d\phi_r(x,v_r)}{dv} = \frac{\partial \phi_r(x,v_r)}{\partial v} + \frac{\partial r(x,v_r)}{\partial v}\frac{\partial \phi_r(x,v_r)}{\partial y} \tag{21}$$

where $\frac{\partial \phi_d(x,v_d)}{\partial y} = \frac{\partial \phi_d(x,v,y)}{\partial y}\Big|_{y=d(x,v)}$ and $\frac{\partial \phi_r(x,v_d)}{\partial y} = \frac{\partial \phi_r(x,v,y)}{\partial y}\Big|_{y=r(x,v)}$

Now from these reduced gradients, we have:

$$\begin{aligned} \| \frac{d\phi_d(x,v_d)}{dx} - \frac{d\phi_r(x,v_r)}{dx} \| &= \| \frac{\partial \phi_d(x,v_d)}{\partial x} - \frac{\partial \phi_r(x,v_r)}{\partial x} \\ &+ \frac{\partial d(x,v_d)}{\partial x}\frac{\partial \phi_d(x,v_d)}{\partial y} - \frac{\partial d(x,v_r)}{\partial x}\frac{\partial \phi_d(x,v_r)}{\partial y} \\ &+ \frac{\partial d(x,v_r)}{\partial x}\frac{\partial \phi_d(x,v_r)}{\partial y} - \frac{\partial r(x,v_r)}{\partial x}\frac{\partial \phi_r(x,v_r)}{\partial y} \| \\ &\le C_3 \| v_d - v_r \| + C_4 \| \frac{\partial d(x,v_r)}{\partial x} - \frac{\partial r(x,v_r)}{\partial x} \| \end{aligned} \tag{22}$$

$$\le C_3\Delta^2 + C_4\Delta \le C_5\Delta \tag{23}$$

where the last inequality follows because the trust region Δ is bounded above. Similarly, it is straightforward to show that:

$$\| \frac{d\phi_d(x,v_d)}{dv} - \frac{d\phi_r(x,v_r)}{dv} \| \le C_6\Delta \tag{24}$$

$$\| \frac{dh_d(x,v_d)}{dx} - \frac{dh_r(x,v_r)}{dx} \| \le C_7\Delta \tag{25}$$

$$\| \frac{dh_d(x,v_d)}{dv} - \frac{dh_r(x,v_r)}{dv} \| \le C_8\Delta \tag{26}$$

Finally, because we can eliminate v along with the equality constraints, $c(x,v)=0$, we can reformulate (9) and (10) as the inequality constrained problems:

$$\min f(x),\ s.t.\ g(x) \le 0 \tag{27}$$

and

$$\min f^r(x),\ s.t.\ g^r(x) \le 0 \tag{28}$$

To demonstrate fully linear properties for the objective and constraint functions, we consider the following functions and reduced gradients:

$$f(x) = \phi_d(x, v_d(x)) \tag{29}$$

$$f^r(x) = \phi_r(x, v_r(x)) \tag{30}$$

$$\frac{df(x)}{dx} = \frac{\partial \phi_d(x,v_d(x))}{\partial x} + \frac{dv_d(x,v_d)}{dx}\frac{\partial \phi_d(x,v_d(x))}{\partial v} \tag{31}$$

$$\frac{df^r(x)}{dx} = \frac{\partial \phi_r(x,v_r(x))}{\partial x} + \frac{dv_r(x,v_r)}{dx}\frac{\partial \phi_r(x,v_r(x))}{\partial v} \tag{32}$$

where $\frac{dv_d(x,v_d)}{dx} = -\nabla_x c_d(x,v_d)(\nabla_v c_d(x,v_d))^{-1}$ and $\frac{dv_r(x,v_r)}{dx} = -\nabla_x c_r(x,v_r)(\nabla_v c_r(x,v_r))^{-1}$.
From these relations it is clear that:

$$\begin{aligned} \| f(x) - f^r(x) \| &= \| \phi_d(x,v_d) - \phi_r(x,v_r) \| \\ &\le C_9 \| v_d - v_r \| \le C_9 \Delta^2 \end{aligned} \tag{33}$$

and

$$\begin{aligned} \| \frac{df(x)}{dx} - \frac{df^r(x)}{dx} \| \le & \| \frac{\partial \phi_d(x,v_d)}{\partial x} - \frac{\partial \phi_r(x,v_r)}{\partial x} \| \\ & + \left\| \frac{dv_d(x,v_d)}{dx} \frac{\partial \phi_d(x,v_d)}{\partial v} - \frac{dv_d(x,v_d)}{dx} \frac{\partial \phi_r(x,v_r)}{\partial v} \right\| \\ & + \left\| \frac{dv_d(x,v_d)}{dx} \frac{\partial \phi_r(x,v_r)}{\partial v} - \frac{dv_d(x,v_r)}{dx} \frac{\partial \phi_r(x,v_r)}{\partial v} \right\| \\ & + \left\| \left(\frac{dv_d(x,v_r)}{dx} - \frac{dv_r(x,v_r)}{dx} \right) \frac{\partial \phi_r(x,v_r)}{\partial v} \right\| \\ \\ \le & C_{10} (\| v_d - v_r \| + \| \frac{dv_d}{dx} - \frac{dv_r}{dx} \|) \le C_{10} \Delta \end{aligned} \tag{34}$$

Following the same analysis for the inequality constraints it is straightforward to show that:

$$\| g(x) - g^r(x) \| \le C_{11} \Delta^2 \tag{35}$$

$$\| \frac{dg(x)}{dx} - \frac{dg^r(x)}{dx} \| \le C_{12} \Delta \tag{36}$$

and fully linear properties hold for the objective and constraint functions in (27) and (28).

4 Extending Algorithm I to Inequality Constrained Problems

We now develop the trust region algorithm based on (27) and (28). To reformulate these problems we form a merit function that includes the objective function and constraint violations. For instance, an exact merit function, whose minimum is equivalent to the solution of (27), is the l_1 penalty function, along with its directional derivative (along direction p):

$$\tilde{\psi}(x) = f(x) + v \sum_{j \in J} \max(g_j(x), 0) \tag{37}$$

$$D_p \tilde{\psi}(x) = \nabla f(x)^T p + v \sum_{j \in J_0} \max(\nabla g_j(x)^T p, 0) + v \sum_{j \in J+} \nabla g_j(x)^T p \tag{38}$$

where $J_0 = \{ j \mid g_j(x) = 0 \}$, $J_+ = \{ j \mid g_j(x) > 0 \}$ and $J_- = \{ j \mid g_j(x) < 0 \}$, $J = J_+ \cup J_- \cup J_0$ and the penalty parameter $v > 0$ is assumed to be bounded above. However, this l_1 penalty function is nonsmooth because of the *max* operator, and this makes it difficult to fit into the framework of the previous section.

On the other hand, we can define a smooth, approximate merit function given by:

$$\psi(x) = f(x) + v\varphi(g(x)) \tag{39}$$

$$\nabla\psi(x) = \nabla f(x) + v\nabla g(x)\varphi'(g(x)) \tag{40}$$

Here we need to assume that $\varphi(g(x))$ has second derivatives that are uniformly Lipschitz continuous. Examples include using the classical quadratic penalty, applying an extended barrier function. (See Srinivasan et al., 2008), or smoothing the max operator in (42). If we apply the smoothed *max* operator, we can make $\psi(x)$ arbitrarily close to $\tilde{\psi}(x)$. A similar trust region framework was developed in Agarwal and Biegler (2011). From the development in the previous section, we can state the following properties:

- From the fully linear property for $f(x), f^r(x), g(x), g^r(x)$ and

$$\begin{aligned}\nabla\psi(x) - \nabla\psi^r(x) &= \nabla f(x) - \nabla f^r(x) + v(\nabla\varphi(g(x)) - \nabla\varphi(g^r(x))) \\ &= \nabla f(x) - \nabla f^r(x) + v(\nabla g(x)\varphi'(g(x)) - \nabla g^r(x)\varphi'(g^r(x))) \\ &= \nabla f(x) - \nabla f^r(x) + v(\nabla g(x)(\varphi'(g(x)) - \varphi'(g^r(x))) \\ &\quad + (\nabla g(x) - \nabla g^r(x))\varphi'(g^r(x)) \\ &= \nabla f(x) - \nabla f^r(x) + O(\Delta^2) + (\nabla g(x) - \nabla g^r(x))\varphi'(g^r(x))\end{aligned}$$

  It is easy to see that the fully linear property holds for the modified l_1 penalty function as well, i.e.:

$$|\psi(x) - \psi^r(x)| \le \kappa_\psi \Delta^2 \tag{41}$$

$$\|\nabla\psi(x) - \nabla\psi^r(x)\| \le \kappa_{g\psi}\Delta \tag{42}$$

- If (27) has a local solution that satisfies KKT conditions and a constraint qualification, then bounded KKT multipliers exist. In this case, it is well known that there exists a finite value of $\bar{w} > 0$ such that the solutions to (27) are equivalent to solutions of the nonsmooth unconstrained problem:

$$\min\ \tilde{\psi}(x) = f(x) + v\sum_{j\in J}\max(0, g_j(x)) \quad \textit{for all } v > \bar{v} > 0 \tag{43}$$

- Similarly, solutions to (28) are equivalent to solutions of the nonsmooth unconstrained problem:

$$\min\ \tilde{\psi}^r(x) = f^r(x) + v\sum_{j\in J}\max(0, g_j^r(x)) \quad \textit{for all } v > \bar{v} > 0 \tag{44}$$

- Solutions x^* of (9), (27) and (43) are equivalent.
- Solutions x^* of (10), (28) and (44) are equivalent.

For nonsmooth problems (43) and (44) gradients of the objective function cannot be used to identify stationary points for the TR algorithm. Instead, Conn, Gould and Toint (2000) use directional derivatives to define the following first-order criticality measures:

$$\chi^r(x) = |\min_{\|p\|\le 1} D_p \tilde{\psi}^r(x)| \tag{45}$$

and

$$\chi(x) = |\min_{\|p\|\le 1} D_p \tilde{\psi}(x)| \tag{46}$$

For the analysis in this study, we do not use these directional derivatives and instead work directly with the smoothed merit functions $\psi(x), \psi^r(x), \nabla\psi(x)$ and $\nabla\psi^r(x)$. The trust region problem analogous to (2) is given by

$$\min_{s\in\Delta_k} \psi^r(x_k + s) = f(x_k + s) + \nu\varphi(g^r(x_k + s)) \tag{47}$$

For the solution of (43) and (47) we apply a fixed penalty parameter ν that is assumed sufficiently large. Moreover, because (44) has an equivalent solution to (10), problem (47) is never solved directly. Instead (10) is solved with the trust region constraint $\| x - x_k \| \le \Delta_k$.

Finally, in the modified trust region algorithm, we also restate the condition for fraction of Cauchy decrease. Instead of (4), we again follow Conn, Gould and Toint (2000) and use:

$$\psi^r(x_k) - \psi^r(x_k + s) \ge \frac{\kappa_{fcd}}{2} \| \nabla\psi^r(x_k) \| \Delta_k \tag{48}$$

The modified algorithm is stated as follows

Algorithm II

1. Choose constants $0 \le \eta_0 \le \eta_1 < 1$ (with $\eta_1 > 0$), $\gamma \in (0,1)$, $\gamma_{inc} > 1$, tolerance ε_t and $\mu > 0$. Choose an initial point x_0 and $\Delta_{\max} > 0$ along with Δ_0 and model $\psi^r(x_0 + s)$

2. At iteration k, if $\nabla\psi^r(x_k) > \varepsilon_t$ then go to Step 3. Else, if $\Delta_k > \mu$, $\nabla\psi^r(x_k)$ or $\nabla\psi^r(x_k + s)$ is not fully linear for $s \in \Delta_k$, then construct a new model $\psi^r(x_k + s)$ that is fully linear on a (possibly smaller) trust region with $\Delta_k := \min\{\mu \| \nabla\psi^r(x_k) \|, \Delta_k\}$. This may require construction of new RMs based on additional evaluations of the ODM.

3. Solve (approximately) the trust region problem (47) and compute the step s_k that provides a fraction of Cauchy decrease (48).
4. Assess the acceptability of the trial point; define

$$\rho_k = \frac{\psi(x_k) - \psi(x_k + s_k)}{\psi^r(x_k) - \psi^r(x_k + s_k)} \tag{49}$$

and consider the following conditions:

- If $\rho_k \ge \eta_1$, set $x_{k+1} = x_k + s_k$ and potentially increase trust region using $\Delta_{k+1} = \min\{\gamma_{inc}\Delta_k, \Delta_{\max}\}$. This is called a *successful iteration*.

- If $\rho_k \geq \eta_0$ and the model is fully linear, set $x_{k+1} = x_k + s_k$ and decrease trust region, $\Delta_{k+1} = \gamma\Delta_k$. This is called an *acceptable iteration*.
- If $\rho_k < \eta_1$ and the model is not fully linear, then set $x_{k+1} = x_k$ and apply model refinement to attempt to make $\psi^r(x)$ fully linear on Δ_k, possibly through additional evaluations of the ODM. This is called a *model improving iteration*.
- If $\rho_k < \eta_0$ and the model is fully linear, then set $x_{k+1} = x_k$ and decrease trust region, $\Delta_{k+1} = \gamma\Delta_k$. This is called an *unsuccessful iteration*.

5. Set $k = k + 1$ and go to Step 2.

Under the assumptions that $\psi(x)$ is bounded below, that the fully linear property holds (or can be made to hold) at x_k and that $f(x)$ and $g(x)$ have Hessians with a bounded norm, it is possible to modify the first order convergence proofs in Conn et al. (2008) to show that $\lim_{k\to\infty} \| \nabla\psi(x_k) \| = 0$.

5 Special Cases from Algorithm II

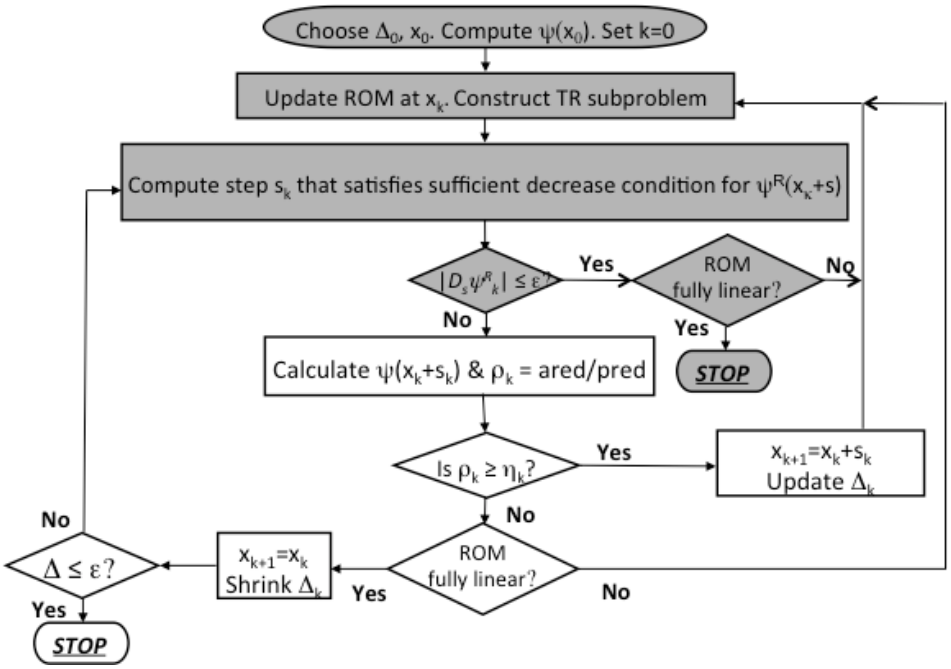


Figure 2. Trust Region Method without Detailed Model Gradients

Algorithm II is sketched in Figure 2. For the flowsheet optimization problem with heterogeneous models, as illustrated by the IGCC process in Figure 1, it is particularly clear that frequent recourse to the original CFD models during the optimization step may be prohibitively expensive. In fact, checking the ODM model at the solution of the RM-based flowsheeting optimization problem may be all that the computing budget allows. To deal with this challenging restriction, we develop RMs with the following characteristics:

- Efficient space-filling experimental designs are used to construct the RMs.
- RMs are poised, i.e., sampling points are chosen and RMs are found that are well-posed and well-conditioned. Kriging models with radial basis functions are a good choice for this.
- The RM-based trust region subproblem (47) is formulated based on the overall flow-sheeting model with an initial large trust region.
- This RM-based subproblem is solved to convergence.

The above characteristics can greatly simplify the algorithm in Figure 2. In fact, since the RM-based model is always fully converged and $\nabla \psi^r(x_k) \leq \varepsilon_t$, only the grayed steps in Figure 2 apply and Steps 3 and 4 in Algorithm II are bypassed. Instead, once the RM-based optimization is determined, the only remaining concern lies in maintaining the fully linear property of the RM.

In the next section, we consider a special case, where the error between the ODM and RM is small in the domain of interest. We call these reduced models ε−exact. Under these conditions, an improved bound can be derived for $\nabla \psi$ as well as the distance between the minimizers of (48) and (49). This leads to a simplication of the algorithm.

5.1 An Improved Bound for $\nabla \psi(x)$

Assume that (8) holds and that the RM and ODM are related by:

$$\| d(u) - r(u) \| \leq \delta \tag{50}$$

for all $u \in \Delta_u \subset U$. This bound can be achieved through extensive pre-computation of RMs using space-filling designs and applying cross-validation to establish the error bound δ. By following an analysis similar to that of Section 3, one can show that this error bound translates to:

$$|\psi(x) - \psi^r(x)| \leq \varepsilon \tag{51}$$

for all $x \in \Delta \subset X$. We term the reduced models that lead to the inequality (56) ε−exact models.

Here we will see that Δ only needs to be large enough to contain a well-poised set of points, x_i, $i = 1, \ldots n_x$. Within some region Δ we choose this set of points and define the matrix:

$$L\delta_L = [(x_1 - x), (x_2 - x), \cdots, (x_{n_x} - x)] \tag{52}$$

where δ_L is the diameter of the set Δ. Also, if we choose a minimal positive basis (see Conn et al., 2009), then $L = I$. From Lipschitz continuity of $\nabla \psi(x)$ and $\nabla \psi^r(x)$, we can write:

$$-\frac{v}{2} \| x_i - x \|^2 \leq (\psi(x_i) - \psi(x)) - \nabla \psi(x)^T (x_i - x) \leq \frac{v}{2} \| x_i - x \|^2 \tag{53}$$

$$-\frac{v_r}{2} \| x_i - x \|^2 \leq (\psi^r(x_i) - \psi^r(x)) - \nabla \psi^r(x)^T (x_i - x) \leq \frac{v_r}{2} \| x_i - x \|^2 \tag{54}$$

with Lipschitz constants $v, v_r > 0$. Noting that:

$$(\psi(x)-\psi(x_i))+\nabla\psi(x)^T(x_i-x)\geq\frac{-\nu}{2}\|x_i-x\|^2 \tag{55}$$

$$(\psi^r(x_i)-\psi^r(x))-\nabla\psi^r(x)^T(x_i-x)\geq\frac{-\nu_r}{2}\|x_i-x\|^2 \tag{56}$$

and from (51), we have:

$$\varepsilon+|\psi(x)-\psi^r(x)|+(\nabla\psi(x)-\nabla\psi^r(x))^T(x_i-x)\geq\frac{-(\nu+\nu_r)}{2}\|x_i-x\|^2 \tag{57}$$

Leading to:

$$2\varepsilon+\frac{(\nu+\nu_r)}{2}\|x_i-x\|^2\geq|(\nabla\psi(x)-\nabla\psi^r(x))^T(x_i-x)| \tag{58}$$

where the absolute value follows because the inequality is symmetric with respect to x and x_i.

By defining $e^T=[1,1,\cdots,1]$ and $e_x^T=[\|x_1-x\|^2,\|x_2-x\|^2,\cdots,\|x_{n_x}-x\|^2]$ we concatenate the above equation over all points x_i, take norms and write:

$$\|2\varepsilon e^T+\frac{(\nu+\nu_r)}{2}e_x^T\|\geq\|(\nabla\psi(x)-\nabla\psi^r(x))^TL\|\delta_L \tag{59}$$

Choosing x_i such that $L=I$, noting that $\|e\|=\sqrt{n_x}$ and $\delta_L^2\geq\|x-x_i\|^2$ gives us:

$$(\frac{2\varepsilon}{\delta_L}+\frac{(\nu+\nu_r)}{2}\delta_L)\sqrt{n_x}\geq\|\nabla\psi(x)-\nabla\psi^r(x)\| \tag{60}$$

Finally, we set $x=\bar{x}$, the solution of (44). This give $\nabla\psi^r(\bar{x})=0$, leading to:

$$(\frac{2\varepsilon}{\delta_L}+\frac{(\nu+\nu_r)}{2}\delta_L)\sqrt{n_x}\geq\|\nabla\psi(\bar{x})\| \tag{61}$$

Since the choice of δ_L is arbitrary, we set it to

$$\delta_L=2(\frac{\varepsilon}{\nu+\nu_r})^{1/2}$$

to minimize the bound on $\nabla\psi(\bar{x})$, which becomes:

$$\|\nabla\psi(\bar{x})\|\leq 2\sqrt{\varepsilon n_x(\nu+\nu_r)} \tag{62}$$

Clearly, when $\varepsilon\to 0$, then $\nabla\psi(\bar{x})\to 0$ and a KKT point for (44) approaches the KKT point of (43)

In addition, we can derive a bound for $x^*-\bar{x}$. Application of Taylor's theorem gives:

$$0 = \nabla\psi(x^*) = \nabla\psi(\bar{x}) + \int_0^1 \nabla^2\psi(x(\tau))(x^* - \bar{x})d\tau$$
$$0 = \nabla\psi(x^*) = \nabla\psi(\bar{x}) + W(x^*, \bar{x})(x^* - \bar{x}) \quad (63)$$

where $x(\tau) = \bar{x} + \tau(x^* - \bar{x})$. Here we note that $\nabla^2\psi(x^*)$ is the reduced Hessian at the optimum of (43) and we make the following assumptions:

- The fully linear property (42) holds over the current trust region Δ.
- For some compact, convex region $\bar{X}$ with $x^*, \bar{x} \in \bar{X}$, $W(x^*, \bar{x})$ is uniformly bounded and positive definite for all $x \in \bar{X}$.

This leads to:

$$\nu_+ \| x^* - \bar{x} \|^2 \geq (x^* - \bar{x})^T W(x^*, \bar{x})(x^* - \bar{x}) \geq \nu_- \| x^* - \bar{x} \|^2 \quad (64)$$

where ν_- and ν_+ are less than the smallest and greater than the largest eigenvalues, respectively, of $W(x^*, \bar{x})$. From (63) we can write

$$(x^* - \bar{x}) = -W(x^*, \bar{x})^{-1}\nabla\psi(\bar{x})$$
$$\| x^* - \bar{x} \| \leq \| W(x^*, \bar{x})^{-1} \| \cdot \| \nabla\psi(\bar{x}) \|$$
$$\leq 2\sqrt{\varepsilon n_x(\nu + \nu_r)}\, \| W(x^*, \bar{x})^{-1} \|$$
$$= 2\sqrt{\varepsilon n_x(\nu + \nu_r)}\, \frac{\kappa(W)}{\| W(x^*, \bar{x}) \|}$$
$$= 2\sqrt{\varepsilon n_x(\nu + \nu_r)}\, \frac{\nu_+}{(\nu_-)^2}$$

where the last three inequalities follow from (63), (64) and the definition of the condition number κ. Assuming that $\nu, \nu_r \leq \nu_+$ gives the bound:

$$\| x^* - \bar{x} \| \leq \frac{(2\nu_+)^{3/2}(\varepsilon n_x)^{1/2}}{(\nu_-)^2} \quad (65)$$

and we note that as $\varepsilon \to 0$, $\bar{x} \to x^*$

From this bound we propose the following refinement strategy for ε−exact models. Moreover, since ν_- and ν_+ are difficult to obtain from the detailed model, we can substitute the eigenvalues ν_-^r and ν_+^r from the reduced Hessian of the *reduced* model. This is likely to be a good approximation for small values of ε.

Algorithm III

1. Choose constants $0 \leq \eta_0 \leq \eta_1 < 1$ (with $\eta_1 > 0$), $\gamma \in (0,1)$, $\gamma_{inc} > 1$, tolerance ε_t and ε_x, and bound parameters ν_+ ν_- for (65). Choose an initial point x_0 and $\Delta_{\max} > 0$ along with Δ_0 and ε_0 to specify exactness of the model.
2. At iteration k, for $\| x - x_k \| \in \Delta_k$, develop a space-filling experimental design and apply a validation procedure, such as cross-validation, to establish an ε−exact set of

reduced models with objective function $\psi^r(x)$ and $|\psi(x)-\psi^r(x)| \le \varepsilon_k$.

3. Solve the trust region problem (47) with solution $\bar{x}$. If $\nabla\psi^r(x) > \varepsilon_t$, go to Step 5.
4. If $\varepsilon_k \le \varepsilon_x$, STOP. Else, set $x_{k+1} = \bar{x}$, $\varepsilon_{k+1} = \gamma\varepsilon_k$, update $\Delta_{k+1} = \{x \mid x - x_{k+1}\}$ using (65) with ε_{k+1}, set k = k + 1 and go to Step 2.
5. Assess the acceptability of the trial point. (*Here the assumption on (65) does not hold and problem (47) terminates successfully at the boundary of* Δ_k. *Since the reduced model is fully linear, we continue instead with the TR steps from Algorithm II.*) Define:

$$\rho_k = \frac{\psi(x_k) - \psi(\bar{x})}{\psi^r(x_k) - \psi^r(\bar{x})} \tag{66}$$

and consider the following conditions:

- If $\rho_k \ge \eta_1$ set $x_{k+1} = \bar{x}$, and potentially increase trust region using $\Delta_{k+1} = \min\{\gamma_{imc}\Delta_k, \Delta_{\max}\}$
- If $\rho_k \ge \eta_0$, set $x_{k+1} = \bar{x}$, and decrease trust region, $\Delta_{k+1} = \gamma\Delta_k$.
- If $\rho_k < \eta_0$, then $x_{k+1} = x_k$, and decrease trust region, $\Delta_{k+1} = \gamma\Delta_k$.

6. Set $k = k + 1$ and go to Step 2.

6 Case Study Examples

In above section it is frequently mentioned the fully linear property that plays a key role in the derivations. In order to guarantee this property, a sophisticated design of experiments (DOE) is important. Briefly speaking, when the radius Δ of a trust region or closed ball is determined, it is necessary to ensure that it contains at least n+1 sampling points within it, on the other hand, the points distribution must satisfy the measure of space-filling. Of course, the smaller the closed ball, the smaller the error in ε-exact models, which means more sampling points are better. To illustrate the importance of the DOE, space-filling and Δ, we give three examples.

Example1. Peak and valley problem

This is an unconstrained minimization problem with two independent variables.

$$\begin{aligned} &\min f(x) = x_1 \exp(-(x_1^2 + x_2^2) + 0.1(x_1^2 + 0.1x_2^2) \\ &s.t. \qquad -2 \le x_1, x_2 \le 3 \end{aligned} \qquad \text{[NLP 1]}$$

The profile of *f(x)* is as shown in Figure 3 in 3-D and its contour. To solve NLP3 with its RM, we developed four ε-exact Kriging models with *Nx* = 27, 48, 75 and 300 sampling points, respectively. The design of experiments (DOE) for the RM is shown in Figure 4 (a)-(d). The leftmost graphic gives the DOE, the middle gives the RM contour and the rightmost shows the approximated 3-D profile by the RM. We then solve NLP3 with its respective RMs, and the results along with the true solution are listed in Table 1.

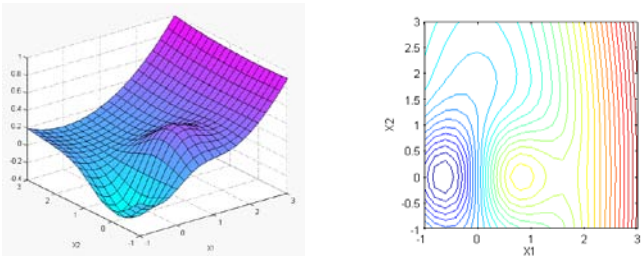
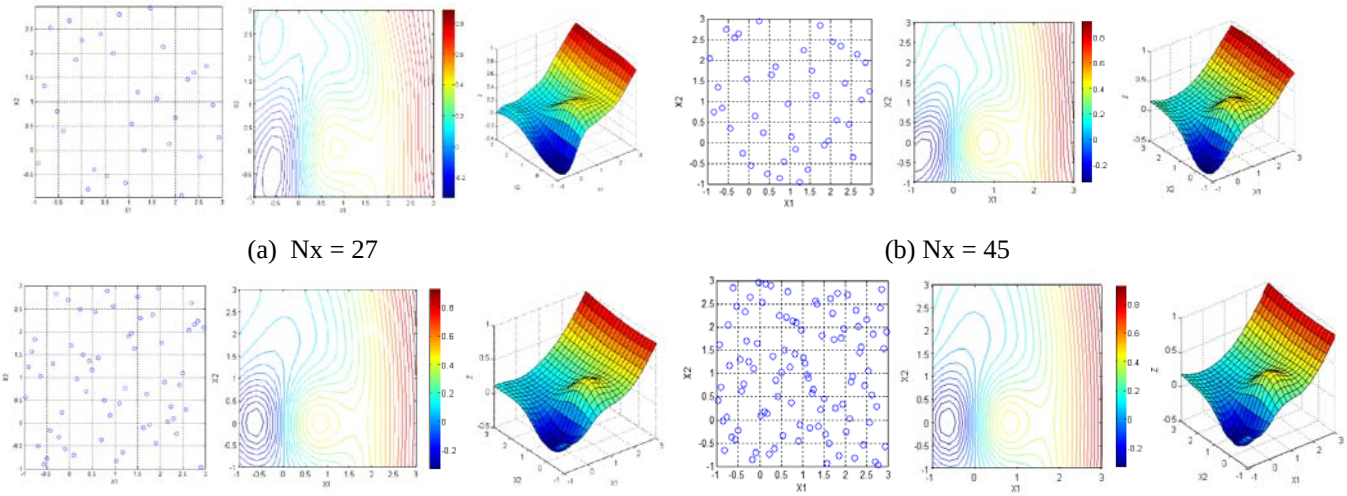


Figure 3. Profiles of NLP1



(a) Nx = 27 (b) Nx = 45

(c) Nx = 75 (d) Nx = 300

Figure 4. RMs of NLP1 with different DOE.
Left–Results of DOE; Middle – contours of RM; Right– 3D profiles

Table 1. Comparison of results with different RMs

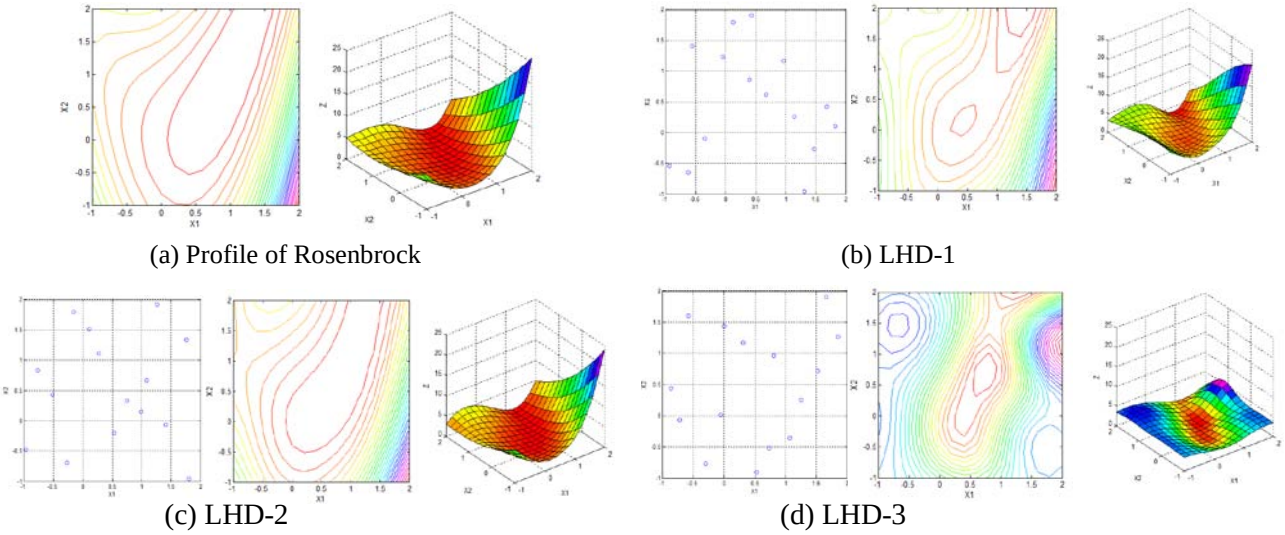
		Objective	X1	X2
True		-0.636127392935813	-0.383961517686256	-0.000000857336992
Nx	**27**	-0.604607912803695	-0.349766822020327	0.015809775403955
	48	-0.634278760540868	-0.376170012473266	0.002851103935254
	75	-0.632737704868186	-0.383469471493864	0.000890245685945
	300	-0.636127448872631	-0.383961555413930	0.000001383330600

The results show that when the error of ε-exact model goes to zero (as *Nx* is increasing), the solution obtained by RM is close to its true optimum, as indicated by (65).

Example 2. Rosenbrock problem

The Rosenbrock function has been used by many authors because its stationary point is located in a flat valley. The problem is the bound constrained minimization of Rosenbrock function

$$\begin{aligned} &\min f(x) = (x_2 - x_1^2)^2 + (1 - x_1)^2 \\ &s.t. \qquad -1 \le x_1, x_2 \le 2 \end{aligned} \qquad \text{[NLP 2]}$$



(a) Profile of Rosenbrock (b) LHD-1

(c) LHD-2 (d) LHD-3

Figure 5. LHD gives different designs even when maximin is used and we pick the best one among 1000 random designs for each of the designs above. (Left – results of LHD; Middle – contours of RMs; Right – 3D profile of RMs)

We intend to use this example to show the importance of space-filling in DOE. Figure 5 gives three LHD results (Figure 5 (b)-(c)) with Nx = 15. It obvious that even if the numbers of the sampling points are the same, Nx =15 the RMs perform significantly differently because of the different distributions of the sampling sets. As reference, we give the true profile of Rosenbrock function as well in Figure 5 (a).

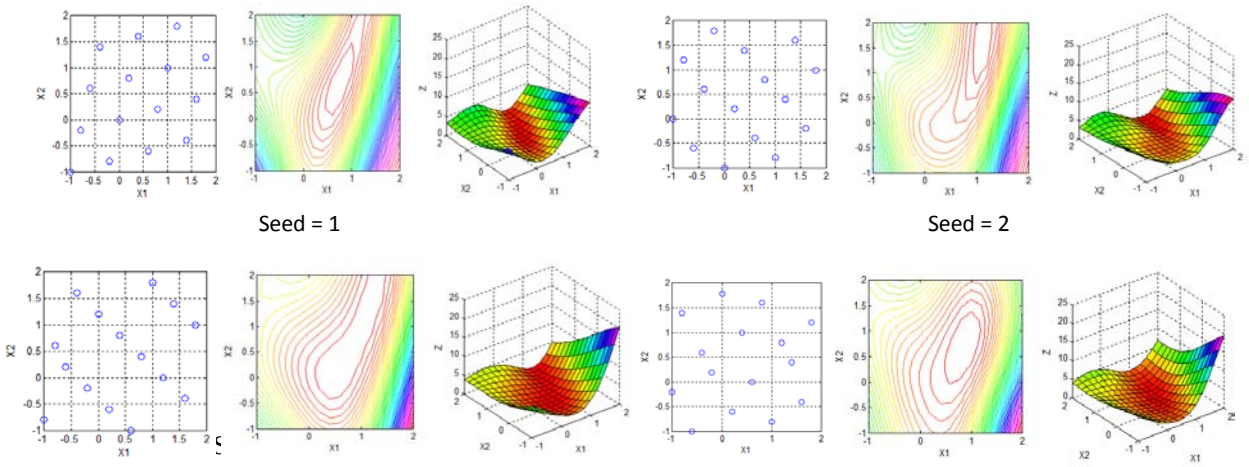


Figure 6. Results of TPLHD with different patterns
Left – results of TPLHD; Middle – contours of RMs; Right – 3D profile of RMs

To improve on LHD we introduce Translation Propagation Latin Hypercube Design (TPLHD) (Viana et al., (2010)). Figure 6 shows the four different patterns used in TPLHD. From the results of TPLHD, we can see that the sampling points are distributed more evenly than that of LHD. In the other word, the results of TPLHD are better in the criterion of space-filling. We can imagine that when

iterative optimization computation is implemented, any new point, x_k on the search path of the optimization has the following characteristics:

> Formulate a closed ball $B_k(x_k, \Delta_k) := \{x \mid x - x_k \le \Delta_k\}$, we can verify that as long as Δ_k is well-designed, B_k can enclose at least n+1 sampling points within it, which guarantees that the RMs possess fully linear property in iteration *k* (Conn et al., 2009).

From Figure 5 (a) we can observe that the minimum of the Rosenbrock is located in a relative flat valley, which means that a small error in RM will lead to larger errors in finding the minimum. The results show that in fact when *Nx* = 48, the solution by RM is close enough to the true optimum. This fact proves that Algorithm III is suitable for ε-exact models, i.e. as long as the RM has small ε, an accurate optimal solution can be still obtained without recourse to its ODM.

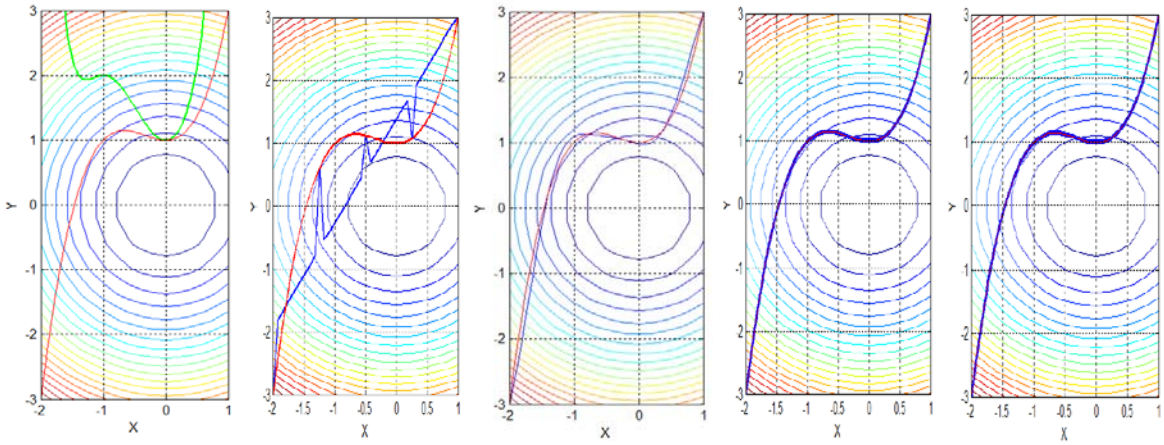
Table 2. Results of Rosenbrock function with its RMs

		X1	X2	Objective
True		1.000000055032763	1.000000075591546	4.217060554804743e-015
Nx	**27**	1.031327347381811	1.056060015084665	-0.010661961379137
	48	0.999869890209556	0.998784689492565	-1.321751617933131e-006
	75	0.998570208150591	0.996507645503141	4.064909972179009e-006
	300	1.000035059566996	1.000081678268593	-8.986803883281880e-007

Example 3. NLP with a cubic constraint

In this example, we only use the constraint, a cubic curve, as the ODM and we choose a Kriging approximation of this constraint as the RM.

$$\begin{aligned} &\min f(x) = x^2 + y^2 \\ &s.t. \quad y = x^3 + x^2 + 1 \qquad \text{[NLP3]} \\ &-1 \le x \le 3, \ -2 \le y \le 3 \end{aligned}$$



(a) (b) (c) (d) (e)

Figure 7. Plots of NLP3 with its RMs of the constraint.

Figure 7 illustrates the profile of the constraint cures and its RMs with different numbers of sampling

points for the Kriging model. Figure 7 (a), shows the profile of objective function (green curve) and the constraint (red) plotted in the contour objective field. From the objective cure we can figure out that there are two local minima, at values of 1.9313 (*@x=-1.28*) and 1.0000 (*@x=0*), respectively. That implies that with different starting points, the NLP3 might converge to different local solutions, which are indicated in Table 3. From (b) to (e), are the profiles of RMs of the constraint with *Nx* =5, 6, 7 and 10, respectively. These plots show how well the RMs, ε -exact models approximate the true constraint. It is seen that when $Nx \geq 7$ ε becomes very small so that no error could be observed from the figures (d) and (e) (red and blue curves are essentially identical). Consequentially, the solutions obtained with the *good* RMs are very close to the ODM. This can be seen in Table 3 that gives two local minima. The results in Table 3 indicate that when the ε -exact models are developed with smaller error (when *Nx* = 7 or 10), they perform just like the ODM.

Table 3. Results of NLP 3 with different ε -exact models

		Starting point	Objective Value	Optimal Variables	
				x	y
ODM		-1.9 2.9	1.9313	-1.2785	0.5448
		-0.9 1.9	1.0000	0.0000	1.0000
ε -exact models					
Nx	5	-1.9 2.9	0.5161	-0.6162	0.3693
		-0.9 1.9	0.5161	-0.6162	0.3693
	6	-1.9 2.9	1.9902	-1.3648	0.3570
		-0.9 1.9	0.9484	0.0516	0.9725
	7	-1.9 2.9	1.9314	-1.2785	0.5448
		-0.9 1.9	1.0000	0.0000	1.0000
	10	-1.9 2.9	1.9313	-1.2785	0.5448
		-0.9 1.9	1.0000	0.0000	1.0000

Example 4. IGCC optimization with CFD-based RMs

We now demonstrate the above approach with the optimization of the advanced IGCC process shown in Figure 1. In addition to the lumped parameter flowsheeting models, the process model contains two CFD models, for the gasifier and combustor. As described in Lang et al. (2009) and Lang et al. (2011), RMs were developed for both CFD models using Latin Hypercube Sampling for the experimental design and PCA-based reduction of the snapshots of their vector fields at each sampling point. For the combustor RM, a Kriging model was developed to map the process inputs to the output streams, while an artificial neural network (ANN) RM was developed for input-output mapping of the gasification unit. Cross-validation on both RMs showed excellent predictive capabilities over the entire input space. Both output temperatures and compositions were typically predicted with relative errors of less that 1%.

Full details of the RM construction and validation of these models are given in Lang et al. (2009). A sample comparison of the RM model is shown for the gasifier in Figure 8. It is notable that sampling points of the CFD combustor model required over 30 CPU min, while the CFD gasifier model required as much as 20 CPU h. On the other hand, each of the resulting RMs could be executed within 1 CPU s.

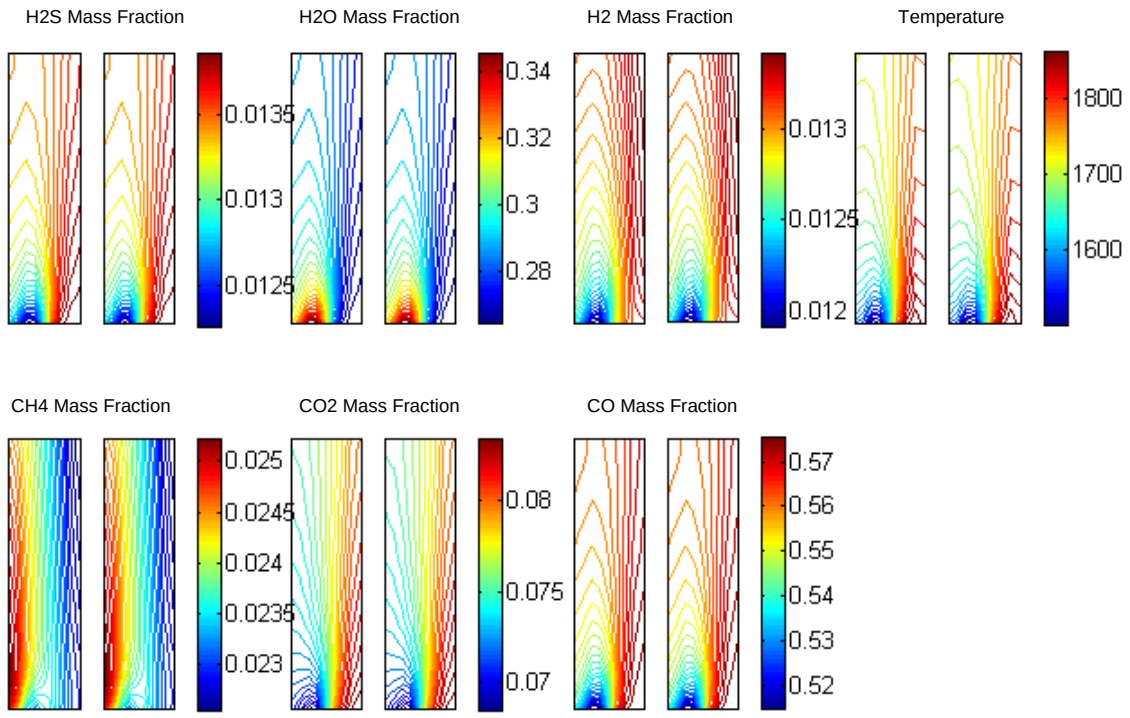


Figure 8. Gasifier vector field with FLUENT result (left) and RM prediction (right)

In our previous work, we integrated CFD based RMs of the gasifier and the combustor within an IGCC power plant flowsheet and implement optimization of the process. The resulting optimization of the power output of the IGCC process with seven independent variables, led to an increase of 23MW or 6.8% (Lang et al. 2011). How is this result related to its true optimum? Following the derivations in above sections, we can be confident that since the RMs of gasifier and the combustor are ε-exact models with smaller ε's, according to (65) the optimal solution of the IGCC flowsheet with RMs are very close to the true optimum, even though this would be very difficult to obtain directly with their CFD models.

7 Conclusions and further works

Starting from a trust region algorithm for derivative free optimization, we extend this algorithm for process flowsheet optimization with multiple scale models. By introducing the exact penalty function, we further extend the algorithm from unconstrained optimization problems to consider constrained problems. With fully linear properties of RMs, we proved that the errors between RMs and their ODM are bounded during the iterative computations which lead to that the optimal solutions with RMs are close to the ODM. Especially, we term ε-exact models with which we can obtain reasonably good approximations to the optimum without recourse to the ODM during the optimization.

With these theoretical derivations, we have confidence that our previous results of IGCC optimization are reliable.

Also we mention that to develop good ε-exact models, sophisticated DOEs are necessary and worth

spending time and effort off-line. Here space-filling DOEs are strongly recommended. With a few simple examples, the performances of our algorithms are illustrated with excellent results.

Finally, we have not implemented any further case studies of flowsheet optimization with multiple scale models; this is beyond the scope of this project, and no resources were provided for this. Nevertheless, this can be investigated by making up a virtual, complicated process in which some units can be replaced by their RMs. With this case, we plan to apply our developed algorithms for further examination.

References

A. Agarwal and L. T. Biegler, "A Trust-region Framework for Constrained Optimization using Reduced Order Modeling," *Optimization and Engineering*, in press (2011)

Conn, A. R., Gould, N. M., Toint, P. L. (2000). *Trust-region Methods, SIAM*. Philadelphia, PA.

Conn, A. R., K. Scheinberg, L. Vicente, *Introduction to Derivative Free Optimization*, SIAM, Philadelphia, 2009

Conn, A. R., K. Scheinberg, L. Vicente, *SIAM J. Opt.*, 20, pp. 387-415, 2008

Lang, Y-D, A. Malacina, L. T. Biegler, S. Munteanu, J. I. Madsen and S. E. Zitney, "Reduced Order Model Based on Principal Component Analysis For Process Simulation and Optimization," Energy and Fuels , 23, 1695–1706 (2009)

Y-D Lang, S. E. Zitney, L. T. Biegler, "Optimization of IGCC Processes with Reduced Order CFD Models," Computers and Chemical Engineering , 35, pp. 1705 – 1717 (2011)

March, A. and Willcox, K., AIAA-2010-9198, AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference (2010)

Srinivasan, B., L. T. Biegler and D. Bonvin, "Tracking the Necessary Conditions of Optimality with Changing Set of Active Constraints using a Barrier-Penalty Function," Computers and Chemical Engineering , 32 (3), p.572-579, (2008)

Wild, S. M., R. G. Regis, C. A. Shoemaker, *SIAM J.Sci. Comp*. 30, pp. 3197-3219 (2008)

Viana, F. A. C., Venter, G. and Balabanov, V., An algorithm for fast optimal Latin hypercube design of experiments. Int. J. Numer. Meth. Engng., 82: 135–156. doi: 10.1002/nme.2750 (2010)